Bank i Kredyt 56(4), 2025, 463-496
DOI: 10.5604/01.3001.0055.2392

BANK I KREDYT
BANK & CREDIT

How reliable are systemic risk measures? Model risk estimates of MES and ΔCoVaR

Aleksandra Pasieczna-Dixit*

Submitted: 19 October 2024. Accepted: 8 January 2025.

Suggested citation: Pasieczna-Dixit A. (2025), How reliable are systemic risk measures? Model risk estimates of MES and ΔCoVaR, *Bank i Kredyt*, 56(4), 463-496, DOI: 10.5604/01.3001.0055.2392.

Abstract

The model risk of two systemic risk measures (SRMs) was quantified for a set of systemically important European banks, using the dispersion of SRM estimates as a proxy. A high model risk was observed, with dispersions of above 65% of the average value, associated with the parametrization error of the Monte Carlo algorithm alone, which has profound implications in the context of systemic risk. Ranking individual banks based on the SRM values was observed to become less dependable due to the high model risk of the SRMs, thus making it difficult for regulators to implement proper policies. Underestimation of the systemic risk of a bank increases the stress within the network, while overestimation of the systemic risk of a bank might lead to undue penalties levied upon the bank. The model risk metric we used additionally allowed us to rank the parameter contributions to the observed model risk.

Keywords: model risk, systemic risk, Marginal Expected Shortfall, Delta Conditional Value at Risk, Monte Carlo

JEL: C63, E44, G21, G28

* Pomeranian Higher School in Starogard Gdański; e-mail: wwwpasiecznapl@gmail.com.

1. Introduction

Systemic risk (SR) is a general concept that is better understood intuitively than formally. Indeed, there is no consensus about its definition. According to Mishkin (1995), SR is defined as the likelihood of a disruptive event that disables financial institutions (FIs) from funding productive investment opportunities. Kaufman (1995) defines SR as the risk of a chain reaction of falling interconnected dominoes. Schwarcz (2008) uses a working definition that combines the risk of an economic shock, the propagation of this shock through the network, and the eventual market-wide increases in cost of capital (or decrease in availability of capital), the last evidenced by financial market-price volatility. Despite the lack of consensus, three components are associated with SR (Kaufman, Scott 2003; Hurd 2016):

1. A shock or a trigger. This refers to events endogenous or exogenous to the financial system, such as an unexpected bankruptcy of a systemically important FI, or a natural catastrophe.

2. Propagation of the shock throughout the financial system. This refers to the spread of the shock/trigger through the entire system, often described as domino, cascade or contagion effects. Two types of channels can exist (Roncoroni et al. 2019) – direct (through inter-FI dependencies such as loans) and indirect (such as exposures to common asset classes). These might amplify shocks through feedback effects present in the network.

3. Potential impact on the wider economy. This refers to the impact felt by the non-financial system part of the global economy in response to the crash of the financial system. There has been an increase in the financialization of the economy (Jajuga 2014), leading to the possibility of more adverse effects to the economy due to a financial crash. These materialize in different forms, such as falls in money supply and stock indices, or significant decreases in economic production and employment (Hurd 2016).

Quantifying SR is not an easy task since each part can contribute in various 'amounts' depending on the financial crisis under study. A result of this is the wide range of computational approaches in literature that aim to measure SR, which we classify in two broad families: network or graph models that build financial networks using propagation channels relying on real or simulated inter-FI information (such as inter-FI loans and exposures to asset classes), and market data based models that infer inter-FI information through correlation measures. This classification is certainly not a strict classification, e.g. one can devise network models that use market data in some parts of the modelling process.

This work uses systemic risk measures (SRMs) that use readily available market data, specifically equity prices and outstanding shares (used to build a market index). Two SRMs are studied in this work: Marginal Expected Shortfall (MES), and Delta Conditional Value at Risk (ΔCoVaR). The MES of an FI studies the impact of a market crash on the FI and is defined as the average price return of the FI when the market falls (Acharya et al. 2010; Brownlees, Engle 2012; Idier, Lamé, Mésonnier 2013). The ΔCoVaR of an FI, on the other hand, looks at how the risk of the market (using the Value at Risk or VaR measure) changes when the FI falls (Adrian, Brunnermeier 2008). Since they both look at the interplay of the market and FIs, they aim to capture the 'too connected to fail' (TCTF) concept. These SRMs are convenient to use, since they rely on market price data, which is often freely available to the wider public, and available at multiple frequencies (such as intra-day, daily, etc.). Higher frequency data can be useful for studying SR, since contagion events can be extremely fast.

Despite the advantages of market-data based SRMs, they have certain limitations. Firstly, they assume that markets are efficient, and so shocks are simulated in the form of FI or market price falls, with little to no regard to the actual cause of the crash (Benoit et al. 2017). Secondly, they infer inter--FI connections through price returns correlations. Given a correlation value, it is difficult to claim, without additional tests, whether the observed value is due to a true inter-FI dependency (where one FI's balance sheet depends on the performance of the other), or an exposure to certain common assets (where the portfolios of both FIs contain similar assets). Finally, Danielsson et al. (2016a) have reported that some of these SRMs have large model risk, making even the significance ranking of FIs unreliable.

Model risk can be thought of as the risk that practitioners are exposed to when using a particular model. Since models are, at best, approximations of observed phenomena, all models have an associated risk when they are used. In the context of SR, this translates to uncertainty about the SRM outcomes. Measurement of SR is of vital importance for regulators and policymakers, who require tools that can properly (and in a timely manner) detect the buildup of stress within the financial system and quantify the contributions of systemically important institutions to the observed stress. The presence of model risk implies that it is harder for regulators to accurately judge the true contribution of a bank to the overall stress. This problem is further compounded when multiple SRMs are combined to capture a wider picture. Indeed, there have been reports in literature that rankings of banks across SRMs can vary quite a lot since they capture different effects (Danielsson et al. 2016b). While some studies have highlighted the variability in SRMs and their implications for regulatory assessments, there is little research focused on quantifying the inherent model risk within a given SRM and understanding its impact on the rankings of systemically important institutions. The study aims to fill this gap by quantifying the model risk of the chosen SRMs and analysing how this propagates to the rankings of the bank. We believe that this can help in the improvement of SRMs and provide useful tools for regulatory frameworks to deal with the inherent uncertainty of SRMs.

Broadly speaking, model risk signifies the uncertainty in model outcomes due to human error in model application (e.g. parametrization, inapplicability) or model shortcomings (e.g. instability, sensitivity). Model risk is thus a consequence of general model construction and uncertainty in finance (Derman 1996; Crouhy, Galai, Mark 1998). Since everything related to a model can be part of the model risk, including data contamination, wrong implementations, badly approximated solutions, software, or hardware bugs, and even the practitioners themselves, it is then not possible to truly mitigate all these risks, especially in the complex domain of finance. As a result, for pragmatic purposes, practitioners look at model risk with more focused points of view. For example, some research works define model risk as inaccuracy arising from estimation errors and uses of incorrect models (Hendricks 1996; Boucher et al. 2014; Glasserman, Xu 2014). Authors might equivalently consider model risk induced by the data--fitting approach used for statistical modelling, namely, the choice of tests for the data and estimation of the model parameters (Sibbertsen, Stahl, Luedtke 2008). Based on various literature sources, and personal experience, we classify model risk sources as:

- Dataset issues

  Poor quality data makes it difficult to extract useful information, whereas incorrect description of datasets might lead to the eventual inappropriate use of the data. Data collection might additionally be subject to biases, such as survivorship bias, where the dataset reflects just the entities that 'survived' until the time of the data collection.
- Data processing issues

  The transformation of data variables, including something as simple as flagging erroneous points or filling missing data, are processes that might potentially increase model risk. These steps are sometimes automated, which, if not managed appropriately, can increase model risk due to human error or the introduction of computer bugs.

- Model construction related issues
  During the abstract development phase (as opposed to the computational implementation phase), certain assumptions and decisions about the model and data usage are made, which can increase model risk if the context changes. Uncertainty in the model parameters, referred to as estimation risk in literature (Klein, Bawa 1976; Lewellen, Shanken 2000), can also contribute to model risk. Estimation risk can be quantified in different ways, such as by building different model variants (Danielsson et al. 2016b; Pasieczna 2021), or by bootstrapping data (Christoffersen, Gonçalves 2004).
- Implementation related issues
  In computational implementations, model risk typically manifests as computer bugs, which are introduced due to multiple factors: human error, faulty assumptions about the data and algorithm, or software and hardware limitations. Model risk also tends to increase with computational complexity, negatively impacting models even in something as simple as spreadsheets. For example, JP Morgan lost over 6 billion dollars due to a small bug in their Excel implementation in 2012 (EuSpRIG 2013; Pollack 2013).
- Model interpretation related issues
  Practitioners use models to help them understand a particular problem, but the models do not capture the entire picture. Relying solely on the models' results can lead to an increase in model risk. Misinterpretation of results can happen due to practitioners using models as 'answer machines' (Wagner, Fisher, Pascual 2010). This is particularly important for regulators and policymakers, as decisions are made based on these results that sometimes affect entire economies.

We distinguish between accuracy, defined as the 'correctness' of the SRM values with respect to a pre-chosen benchmark or property, and model risk, defined as the variation in SRM estimates. Additionally, we focus on the model parametrization error, sometimes referred to as estimation risk (Klein, Bawa 1976; Jorion 1996; Christoffersen, Gonçalves 2004). We quantify this risk by building SRM estimates under different parameter values and measuring the variation of these estimates. The remainder of the paper is organized as follows. In the following section, we describe: (i) the dataset used, (ii) the construction of the market index, (iii) the SRMs analysed, (iv) the model variants for generating the SRM estimates along with the Monte Carlo (MC) algorithm used, and (v) the model risk metric used. Section 3 contains the model risk results and the ranking of parameter contributions to the model risk. We end the work with concluding remarks.

2. Data and methodology

2.1. Data

The FIs chosen in this study are a subset of a list of FIs directly supervised by the European Central Bank (ECB). The ECB considers them to be systemically important and regularly updates this list.[1]

[1] European Central Bank, List of supervised banks, https://www.bankingsupervision.europa.eu/banking/list/html/index.en.html, retrieved 22 August 2021.

These FIs correspond to countries with the euro as their official currency, or countries whose currencies are pegged to the euro. The criteria for their choice are described on the ECB's Banking Supervision website.

The data collection was done in August 2021, at which time the ECB supervised 114 FIs, of which 47 were kept for study. The remaining FIs were rejected for one of the following reasons: unavailability of data in Bloomberg, unlisted (or pending listing) on the market exchanges, a private company, or having been acquired by an included FI. The list is provided in Table 1. The data consisted of daily close prices and daily outstanding shares of the chosen FIs and was obtained using the Bloomberg terminal. The daily outstanding shares data was necessary to compute the market capitalization of the FIs for building a reference market index (see Section 2.2). Whenever possible, the time series for each FI began on the first of January 2000.

2.2. Index building

The SRMs considered in this study require a reference market index. Typically, an appropriate existing index, such as Euro Stoxx, or Euro Stoxx Banks, is used, which represents the actual European 'market'. However, since we are not evaluating the SRMs themselves, but the model risk associated with their estimation, we opted to construct a simple market-capitalization weighted index using only the considered FIs, since an external market index may have exogenous (exogenous to the list of FIs) effects that affect the model risk estimation. Conceptually, the self-built index represents the amount of wealth generated by these FIs, and the index movements are solely due to the movements of the considered FIs. Mathematically,

$$\text{index}_d = \frac{\sum_{i \in \text{comp}_d} \text{mcap}_i(d)}{\text{divisor}_d} \tag{1}$$

where index_d, divisor_d, $\text{mcap}_i(d)$, and comp_d are the values of the market index, the divisor, the market capitalization of FI *i*, and the market composition respectively, on the day *d*; the market capitalization of the FI is the product of its market price and outstanding shares.

Since the index represents the total wealth of the system, market prices and outstanding shares were forward filled, i.e. missing data were filled with the last known values. The divisor was updated whenever a new FI entered the market as described in eq. (2), and was used from the following simulation day to ensure causal use of data. The starting value of the divisor was chosen as described in eq. (3), so that the index began with a value of 100 points.

$$\text{divisor}_{d+1} = \text{divisor}_d \frac{\sum_{i \in \text{comp}_d} \text{mcap}_i(d)}{\sum_{i \in comp_{d-1}} \text{mcap}_i(d)} \tag{2}$$

$$\text{divisor}_{d=2} = \frac{1}{100} \sum_{i \in \text{comp}_{d=1}} \text{mcap}_i(d=1) \tag{3}$$

The final market index is shown in the top sub-plot of Figure 1. We expect this index to have survivorship bias. For example, an FI that was supervised before 2014, but not supervised after, will not be present in the dataset, since the supervised list used was that for August 2021. The prices of this FI would not have been downloaded, thus it was missing in our simulations. However, we argue that if the information about the FI prices was significant at that time, it would already have been incorporated into the price movements of the other existing FIs through the inter-FI correlations – assuming that the price movements reflect all available information.

To verify the quality of the self-built index, we compared it with a free-float market-capitalization weighted European-banks index, the Euro Stoxx Banks (SX7E) index,[2] shown in the bottom sub-plot of Figure 1. The SX7E data was downloaded from investing.com and began on 28 December 2012. Only the overlapping period is shown in the sub-plot. We see that the self-constructed market index has a visual behaviour remarkably similar to the SX7E index, with the survivorship bias manifesting as mild optimism (our index slightly outperforms the SX7E). Since the correlations between the log-returns of the self-built and SX7E indices are 97.76%, and since we have limited data for the SX7E index, we use the self-built market index in this work.

2.3. Systemic risk measures

We analysed the model risk of two widely used SRMs:

A. Marginal Expected Shortfall (MES)

MES is defined as the average return of the FI when the market returns are in their left tail (Acharya et al. 2010; Brownlees, Engle 2012; Idier, Lamé, Mésonnier 2013):

$$\text{MES}_{i,t}(\alpha) = E\left(R_{i,t} \mid R_{m,t} \le \text{VaR}_{m,t}(\alpha)\right) \tag{4}$$

where:

$R_{i,t}$ and $R_{m,t}$ represent the price returns of the FI i and the market m respectively at time t,

$E(x)$ is the expectation value of x,

$\text{VaR}_{m,t}(\alpha)$ is the VaR of the market at confidence level α, and is defined as the maximum possible loss, whose probability is within a pre-defined confidence level over a predefined time horizon (Hendricks 1996; Holton 2014; Pasieczna 2019).

In our MC setup, the MES of a bank is computed as the average simulated returns of the bank over the selected MC iterations, where the market's simulated returns are below the simulated market VaR.

B. Delta Conditional Value at Risk (ΔCoVaR)

The ΔCoVaR studies the FI's contribution to overall SR (Adrian, Brunnermeier 2008; Castro, Ferrari 2014), and is defined as:

$$\Delta\text{CoVaR}_{i,t}(\alpha) = \text{CoVaR}_t^{\left[m|R_{i,t}=\text{VaR}_{i,t}(\alpha)\right]}(\alpha) - \text{CoVaR}_t^{\left[m|R_{i,t}=\text{VaR}_{i,t}(0.5)\right]}(\alpha) \tag{5}$$

[2] EURO STOXX Banks, https://stoxx.com/index/sx7e/, retrieved 27 August 2022.

$\mathrm{CoVaR}_t^{[m|R_{i,t}=\mathrm{VaR}_{i,t}(\beta)]}(\alpha)$ is defined as the VaR of the market m at confidence level α, when the returns of the FI i are at their VaR (confidence level β). $\beta = 0.5$ implies median returns at the FI ('normal' functioning). Thus, ΔCoVaR quantifies the change in market risk (market VaR) as the FI crashes. We use a slightly different approach to a quantile regression (Bianchi, Sorrentino 2020) to estimate it, where we compute the market VaR over a certain range (1% = ±0.5% of the total MC iterations) of simulated market returns around the required quantile (i.e., $\beta \pm 0.005$) of the FI's simulated returns.

The MES studies the reaction of the FI to market falls, whereas the ΔCoVaR quantifies the reaction of the market to FI falls. We expect them to behave differently, except in two extreme cases. First, if one FI is so big that it alone dominates the market index, then both SRMs will get correlated to each other. Second, if all FIs in the market index have extremely large exposures to a common factor (strong correlations between FIs), then the market index picks up the common exposure, and both SRMs will get correlated to each other. In our case, the market index consists of FIs whose relative sizes are less extreme with a maximum of 11% (left image in Figure 2), and the inter-FI correlations, while positive with an average of 39%, are not very high (right image). So, we expect these SRMs to act relatively independently of each other.

2.4. Model variants and Monte Carlo algorithm

In this work, we quantify the model risk of SRMs due to the parametrization error of the MC distribution used to simulate the uncertainty in price returns (defined as daily natural log price changes). The MC distribution is used to simulate different potential future trajectories of individual FIs and the market index and is estimated from historical data with multiple parameters. By altering these parameters, we simulate the uncertainty in the MC distribution directly. This proposed method requires significantly fewer trials to get a model risk estimate compared to a data bootstrapping approach. Four parameters were altered in this work:

A. Inter-FI correlations:
  a) independent: each FI's price returns are treated independently, i.e. the correlation matrix is fully diagonal; the interaction between them is incorporated through a correlation between the returns of one FI and the market index and simulated with a bivariate process (Brownlees, Engle 2012);
  b) multivariate: the full correlation matrix is used to obtain simulated price returns for each FI, and the market index is built from the corresponding simulated prices.

B. Distribution type:
  a) normal: a symmetric bell-shaped distribution, characterized by two parameters: the inter-FI covariance matrix $\mathbf{\Sigma}$ and the mean $\boldsymbol{\mu}$ in the multivariate case, and the standard deviation $\boldsymbol{\sigma}$ and the mean $\boldsymbol{\mu}$ in the independent case;
  b) Student's-t: a symmetric bell-shaped distribution like the normal distribution but allows for heavier tails through an additional degree of freedom parameter $\boldsymbol{\nu}$, which is estimated for each FI (and the market in the independent case).

C. Window type:
  a) rolling: simple moving statistical estimates over a certain window length with equal weights to all days;

b) exponentially weighted moving (EWM): moving statistical estimates which give more importance to recent history; here, the centre of mass (com) parameter was used to define the EWM timescale, which decreases the weight by a factor of $\frac{\text{com}}{1+\text{com}}$ per day; the weight of a data point at is about 37% of the weight of a data point on the first day, and we expect the weights to be negligible by 5 xcom.

D. Historical period ('days' refers to trading days):

a) 125 days: approximately 6 months of data;
b) 250 days: approximately 1 year of data;
c) 500 days: approximately 2 years of data.

Within the MC field, there are two main viewpoints on whether to consider the distribution choice (normal/Student's-t, independent/multivariate) in the data generating process (DGP) as parameters to the model or not. In classical statistical modelling, the choice is typically viewed as part of the model specification, with parameters estimated within that chosen distribution family (Robert, Casella 2013; Gelman et al. 1995). In contrast, in Bayesian approaches, by placing priors over different distribution families (or the entire space of distributions), the distribution choice itself acts as a parameter (Hoeting et al. 1999). For example, Green introduced the Reversible Jump Markov Chain Monte Carlo techniques, in which a model indicator was included so that one could sample not just the distribution parameters, but the distribution family itself, such as jumping from a normal distribution with two parameters to a Student's-t with three parameters (Green 1995). The first perspective thus fixes the distribution family and focuses on estimating parameters within it, whereas the latter incorporates uncertainty about the distribution itself in the modelling process.

Within the context of SRMs, researchers tend to compare the estimates from different distributions (Danielsson et al. 2016b; Löffler, Raupach 2018), which is why we align ourselves to the second viewpoint, which considers the distribution choice as a parameter. To obtain a complete picture, it is instructive to estimate model risk from both viewpoints. Therefore, the next section presents the model risk results for each perspective.

In addition to the altered parameters, various choices were further made for purposes such as ease of computation, numerical stability, or 'realistic' results. We highlight some of them, since they stabilize the SRM estimates, and lower the model risk.

The degrees of freedom parameter $\mathbf{v}$ was estimated with the rolling approach only, as the scipy implementation (Virtanen et al. 2020) used here does not support weights for individual observations while fitting non-normal distribution parameters. This causes $\mathbf{v}$ to be static across the window type parameter, leading to a smaller variation in the SRM estimates and a consequent drop in model risk estimates. Furthermore, the values of $\mathbf{v}$ were lower bound to 5. $\mathbf{v} < 5$ implies an extremely heavy-tailed distribution, and as the Student's-t distribution is symmetric, we observed unrealistically large positive price returns. As the distributions are not allowed to be arbitrarily heavy tailed, we expect a drop in model risk estimates.

In terms of data processing, missing prices were dropped while computing price returns to minimize the number of zero-price changes. Multiple zero-price changes reduce the standard deviation and $\mathbf{v}$ parameter, leading to the distribution appearing less risky but more heavy-tailed. However, dropping data points leads to less reliable statistical estimates. As a compromise, the FI's mean, standard deviation and $\mathbf{v}$ were estimated when there was data of at least 80% of the historical period parameter, and the inter-FI correlation was estimated when there was data of at least 60% (a maximum of 20% data missing for each FI). As the SRM estimates become more stable, we expect lower model risk estimates.

We now describe the MC algorithm used to generate the SRM estimates for the different model variants. For each day d of the simulation period, the MES and ΔCoVaR at confidence level α for day $d+1$ were calculated as follows (quantities with a hat $\hat{x}$ imply simulated quantities):

1. Estimate the empirical mean $\mu_{i/m,d}$, standard deviation $\sigma_{i/m,\ d}$ and the degrees of freedom parameter $\nu_{i/m,\ d}$ for each FI i and the market m, using the estimation window type and historical period parameters. Given the inter-FI dependencies parameter, either compute the correlations $\rho_{im,\ d}$ of each FI with the market (independent) or the full inter-FI correlation matrix $\boldsymbol{\Sigma}_d$ (multivariate).

2. With the estimated statistical quantities, generate $N = 20000$ random returns (index n indicates the MC iteration) for each FI and the market for day $d+1$, using one of the following approaches (inter-FI dependencies parameter):

a) independent: generate bivariate random numbers (Brownlees, Engle 2012):

$$r^n_{m,d+1} = \sigma_{m,d}\varepsilon^n_{m,d} \tag{6}$$

$$r^n_{i,d+1} = \sigma_{i,d}\rho_{im,d}\varepsilon^n_{m,d} + \sigma_{i,d}\sqrt{1-\rho^2_{im,d}}\,\varepsilon^n_{i,d} \tag{7}$$

where $\varepsilon^n_{m/i,d}$ represent the random numbers (n^{th} MC iteration) generated using a standard normal/Student's-t distribution corresponding to the market and the FI; the market and FI returns $r^n_{m/i,d+1}$ were simulated using the $\varepsilon^n_{m,d}$; simulated prices for the FIs and the market were obtained as $p^n_{i/m,d+1} = p_{i/m,d}\exp\left(r^n_{i/m,d+1}\right)$;

b) multivariate: as the empirical inter-FI correlation matrix is used, the returns $r^n_{i,d+1}$ were generated for all FIs simultaneously with the multivariate Gaussian or Student's-t distributions; these simulated FI returns were converted to simulated prices as $p^n_{i,d+1} = p_{i,d}\exp\left(r^n_{i,d+1}\right)$, and then aggregated to a simulated market index $p^n_{m,d+1}$; the simulated market returns were computed as $r^n_{m,d+1} = \log\left(p^n_{m,d+1}\right) - \log\left(p_{m,d}\right)$.

3. Using the simulated returns, the MES and ΔCoVaR were estimated for the risk condition α, defined as 95% or 99% in this work, as:

a) MES: estimate the VaR of the market as the α^{th} quantile from the N simulated market returns and average the simulated returns of the FIs over iterations when the simulated market returns were below this VaR;

b) ΔCoVaR: first compute the stressed VaR as the VaR of the simulated market returns at confidence level α over MC iterations where the simulated FI returns were between their $(\alpha+\varepsilon)^{\text{th}}$ and $(\alpha-\varepsilon)^{\text{th}}$. Then compute the unstressed CoVaR as the VaR of the simulated market returns at confidence level α where the simulated FI returns were between their $(0.5+\varepsilon)^{\text{th}}$ and $(0.5-\varepsilon)^{\text{th}}$ quantiles. The difference between the stressed and unstressed CoVaRs is the ΔCoVaR. ε was set to 0.005, allowing us to have 1% of simulated returns (200 iterations) for the estimation of the stressed and unstressed CoVaRs.

Four parameters are altered in the MC algorithm, and thus, we have 24 estimates of the MES and ΔCoVaR per FI per day per risk condition ($\alpha \in \{95\%, 99\%\}$). As mentioned earlier in this subsection, we also split the estimates per DGP, which gives us six estimates of the MES and ΔCoVaR per DGP per FI per day per risk condition.

2.5. Model risk

Here, model risk implies the dispersion of SRM estimates, measured by the 'spread,' defined as the difference between the maximum and minimum values normalized by the average value:

$$\text{Model Risk}_{\text{spread}} = \frac{\text{SRM}_{\max} - \text{SRM}_{\min}}{\text{SRM}_{\text{avg}}} \tag{8}$$

This metric additionally allows one to rank parameter contributions to the total model risk. This can be done by first computing the average of the model risk values obtained by fixing an individual parameter to its values, and then by measuring the drop with respect to the total model risk. A higher drop implies a larger contribution of a parameter, since the 'unexplained' part of the model risk is due to the variation of this parameter. This type of analysis is referred to as sensitivity analysis and is used to devise modelling schemes robust to parameter choices. For comparison of parameter contributions, an important property of a model risk metric is that the model risk of a subsample is less than or equal to the total model risk. The spread metric satisfies this property if the denominator in eq. (8) contains the average of the full sample.

Other than the spread, one can use the ratio of the maximum to minimum values as a model risk metric (Danielsson et al. 2016b; Pasieczna 2021; Pasieczna, Szydłowska 2021), which is equivalent to the spread, up to a shifting and scaling transformation:

$$\text{Model Risk}_{\text{spread}} = \frac{\text{SRM}_{\max} - \text{SRM}_{\min}}{\text{SRM}_{\text{avg}}} = \frac{\text{SRM}_{\min}}{\text{SRM}_{\text{avg}}}\left(\frac{\text{SRM}_{\max}}{\text{SRM}_{\min}} - 1\right) = \frac{\text{SRM}_{\min}}{\text{SRM}_{\text{avg}}}\left(\text{Model Risk}_{\text{ratio}} - 1\right) \tag{9}$$

The ratio satisfies the property that the ratio of the subsample is smaller than or equal to the ratio of the full sample. The main advantage of the spread is that it is slightly better behaved if the smallest estimates are close to zero, in which cases the ratio values get arbitrarily large.

A natural extension of the spread is the standard deviation over the average, which is more robust to outliers (single extreme maximum or minimum values). However, we would need more SRM estimates, which is not the case here, where we work with 24 estimates. In addition, the property of the subsample having a standard deviation not greater than the full sample does not always hold, and so ranking parameter contributions is not easy.

As the spread is more robust than the ratio to smaller values, can work with fewer estimates, and has the property that subsamples cannot have model risk larger than that of the full sample, we use the spread here.

While the spread metric allows one to quantify the differences between the estimates for a single FI, regulators tend to look at the relative contributions of the banks based on their SRM estimates before clustering them into risk buckets. Indeed, the Financial Stability Board assigns systemically important banks into six risk buckets (bucket-0 to bucket-5 as of December 2024), with FIs in bucket-0 having no requirement of an extra capital buffer, and FIs in bucket-5 (currently empty) having an extra capital buffer requirement of 3.5%.[3] As a proxy of a regulatory framework, we performed a ranking of the MES and ΔCoVaR estimates across all FIs for each MC parameter combination per day. We then bucketed

[3] Financial Stability Board, List of Global Systemically Important Banks (G-SIBs), https://www.fsb.org/2024/11/2024-list-of-global-systemically-important-banks-g-sibs/, retrieved December 2024.

the banks into six equally sized buckets B0–B5, with B0 containing the lowest-risk FIs, and B5 containing the highest-risk FIs. Thus, for each FI, per day per SRM, we have 24 estimates of its risk bucket.

To quantify the model risk of the bucketing, we used two metrics: percentage agreement, and the Krippendorff's alpha (Krippendorff 2019; Castro 2017). These allows us to treat the parameter combinations as raters and then quantify the level of agreement between the assignments of the banks to their risk buckets. The difference between the two metrics is that Krippendorff's alpha corrects for chance agreement, i.e. raters might agree 'accidentally', whereas the percentage agreement does not. Additionally, the percentage agreement only looks at whether the assignments match and does not have a notion of the distance between the buckets. This implies that the metric penalizes B0/B1 in the same way as B0/B5. The Krippendorff's alpha, on the other hand, supports ordinal-rated data, like in our case. Both metrics are estimated daily for both SRMs.

3. Results and discussion

The average spread results for the MES and ΔCoVaR at both confidence levels are provided in Table 2. Model risk estimates with fixed individual parameters are also provided so that parameter contributions can be ranked. The spread values for the MES (ΔCoVaR) at 95% and 99% confidence levels were 0.677 (0.854) and 1.137 (1.850) respectively, indicating that the distance between the maximum and minimum was nearly 68% (85%) of the average in the low-risk condition and over 110% (180%) of the average in the high-risk condition. These values signify high model risk for both SRMs, since with a spread value as large as the average, the individual estimates lie anywhere from half the average to 1.5 times the average!

High model risk values raise questions on the model validity, particularly when single estimates are generated. Consider the example when the ΔCoVaR at 95% for a given FI is 0.10% in percentage price change units. Assuming this is an average estimate, with a symmetric spread value of 85% (±42.5%), the 'real' ΔCoVaR can be anywhere between 0.05% (0.1 × 0.575) and 0.1425% (0.1 × 1.425), a factor of about 3 between the extremums. Using these results to identify systemically important FIs amplifies the model risk due to the uncertainty across FIs. While one might want to consider the distribution of SRM estimates to generate a band based on confidence intervals that reflects the majority distribution of the estimates, the high model risk might still make the results unreliable.

To determine the contribution of the parameters to the overall model risk, we compute the drop in model risk by fixing parameter values, denoted by the column $\Delta_P^{x\%}$ in the table, along with the average drop over all parameter values for a given parameter, denoted by the column $\Delta_P^{x\%}$. A larger value of the average drop for a particular parameter implies a higher contribution, since changing this parameter was the factor in the drop of the total observed model risk.

When the inter-FI dependencies parameter is fixed, the spread values for the MES (ΔCoVaR) at 95% and 99% confidence levels changed on average by -0.096 (-0.154) and -0.255 (-0.519) respectively. The multivariate approach has slightly lower model risk than the univariate (independent) approach, indicating that a mean-field type approach contributes a bit more to model variability.

Changing the distribution type parameter changes the spread values for the MES (ΔCoVaR) at 95% and 99% confidence levels on average by -0.189 (-0.267) and -0.373 (-0.678) respectively. Fixing the parameter to the normal distribution leads to a stronger reduction than fixing the parameter to the Student's-t distribution, indicating an increased sensitivity of both SRMs to the modelling of the tail behaviour.

The window type parameter appears to contribute less to the model risk than the distribution type parameter for both SRMs. The spread changed for the MES (ΔCoVaR) at 95% and 99% confidence levels on average by -0.136 (-0.139) and -0.190 (-0.240) respectively. The difference in the reduction is not identical across the parameter values, indicating a small sensitivity to the shape of the window function used for estimating statistics of historical data.

The largest contribution to the total model risk appears to be from the historical period parameter. On average, the spread for the MES (ΔCoVaR) at 95% and 99% confidence levels is reduced by -0.289 (-0.294) and -0.432 (-0.566) respectively. Estimates with a given historical period are more similar to each other than those with a different value of the parameter, and when the parameter values are altered, the model risk increases. This parameter is linked to the memory of past events and reactivity to new data, with smaller time scales being quicker in forgetting past data and reacting to new data. When estimates with different time scales are compared, the model risk is perceived to be increased.

We plot the temporal evolution of the spread for the MES at 95% and 99% confidence levels for all FIs in Figure 3, along with the average and median. The black dashed line at zero represents the 'ideal' (desired) situation of no model risk. An immediate observation can be made that the spread never reaches the ideal values for any day for any FI. We also notice the model risk for the MES at 99% confidence level tends to be higher than that for the 95% confidence level, which can be attributed to the fact that estimates under the high-risk condition vary a bit more due to the dependence on the tails. The model risk is observed to increase typically just after high volatile periods (e.g. 2008–2009, 2020). Estimates can have large variations when markets get volatile because of the different reaction times, memory, or tailedness. As volatile periods end, the model risk increases due to this variation.

The temporal evolution of the spread for the ΔCoVaR at 95% and 99% confidence levels for all FIs is provided in Figure 4, along with the average and median. Like with the MES, the three observations hold: (i) the spread never reaches the ideal values for any day for any FI, (ii) the model risk for the ΔCoVaR at 99% confidence level tends to be higher than that for the 95% confidence level, and (iii) increases in model risk typically just after volatile periods (e.g. 2008–2009, 2020).

Thus, the model risk behaviour is similar for both SRMs, with the ΔCoVaR having a higher model risk than the MES, hinting at the increase of model risk with model complexity. The fact that model risk is always present has profound implications for practitioners, who must adapt to deal with it. Furthermore, the individual FI results for both SRMs indicate that there can be very extreme values of the spread compared to the average curve. As an example, in the period 2018, the individual spread values for the MES at 99% confidence level can reach nearly 3 (average of 1.5). Since the spread is left bounded (minimum value is zero), we can infer that the model risk is right skewed with fat right tails. The skewness values for the spread per day (skewness measured across the FIs) for the MES (ΔCoVaR) at 95% and 99% confidence levels were 0.8367 (0.8304) and 0.9057 (0.9094) respectively. The right-skewed nature means that the risk associated with model failure is large, and within the context of SR, can be very damaging for investors, firms, and the macro-economy.

Table 2 contains the results where the distribution choice is treated as a parameter within the algorithm. As mentioned previously (Section 2.4), there is another viewpoint where the data-generating process (DGP) is a model specification, and the distribution parameters (means, variances, etc.) are estimated from the sampling parameters (here, the historical period, and the window type). Based on this viewpoint, we thus have 4 models corresponding to the four DGPs (normal/Student's-t, independent/multivariate). Table 3 and Table 4 contain the spread results for MES and ΔCoVaR SRMs respectively for the four models at both confidence levels, along with the drops in model risk when the sampling parameters were fixed.

From Table 3, we see that the spread for the MES estimated with the normal model with independent FIs at 95% (99%) confidence level has an overall model risk of 0.354 (0.455). When fixing the window type parameter, the model risk drops by an average of 0.115 (0.148). By fixing the historical period parameter, the spread drops by an average of 0.231 (0.294). The overall spread for the MES estimated with the normal multivariate model at 95% (99%) was 0.359 (0.460). The model risk dropped on average by 0.118 (0.150) when fixing the window type parameter, and by 0.234 (0.298) when fixing the historical period parameter. In the case of the MES computed with the Student's-t distribution with independent FIs at 95% (99%) confidence level, the spread was 0.501 (0.791), which dropped on average by 0.146 (0.214) when fixing the window type parameter, and by 0.346 (0.563) when fixing the historical period parameter. Finally, the spread for the MES with the Student's-t multivariate distribution at 95% (99%) confidence level was 0.422 (0.549), which dropped on average by 0.133 (0.173) when fixing the window type parameter, and by 0.281 (0.361) when fixing the historical period parameter.

By analysing the results in Table 4, we see that the spread for the ΔCoVaR estimated with the normal model with independent FIs at 95% (99%) confidence level has an overall model risk of 0.360 (0.534). When fixing the window type parameter, the model risk drops by an average of 0.117 (0.174). By fixing the historical period parameter, the spread drops by an average of 0.221 (0.325). The overall spread for the ΔCoVaR estimated with the normal multivariate model at 95% (99%) was 0.372 (0.547). The model risk dropped on average by 0.120 (0.177) when fixing the window type parameter, and by 0.230 (0.335) when fixing the historical period parameter. In the case of the ΔCoVaR computed with the Student's-t distribution with independent FIs at 95% (99%) confidence level, the spread was 0.559 (1.155), which dropped on average by 0.159 (0.294) when fixing the window type parameter, and by 0.378 (0.824) when fixing the historical period parameter. Finally, the spread for the ΔCoVaR with the Student's-t multivariate distribution at 95% (99%) confidence level was 0.485 (0.725), which dropped on average by 0.152 (0.232) when fixing the window type parameter, and by 0.308 (0.451) when fixing the historical period parameter.

The general conclusions hold for the model risk even when we consider the distribution choice as a model specification. We find that the model risk for the SRMs under the higher-risk conditions (99% confidence level) is higher than that for the lower-risk condition (95% confidence level). Additionally, the ΔCoVaR tends to have higher model risk than the MES, indicating that model risk increases with model complexity. When comparing the four models among themselves, we find that the models built on Student's-t distributions tend to have higher model risk than those built with normal distributions. This is also linked to model complexity, as the Student's-t has an additional parameter (degrees of freedom) compared to the normal distribution. Finally, we observed that when using the Student's-t distribution, having a multivariate distribution had a reduced model risk compared to having independent FIs connected through the market index. This indicates a potential instability in the bivariate process, even though both approaches are equivalent, since the market index does not have any exogenous effects. This difference was not observed when using the normal distribution, leading us to conclude that the potential instability might be in the modelling of the tails.

The spread metric can also be used to judge the parameter contributions to model risk. Before ranking the parameter contributions, it helps to analyse one important property of parameter contributions measured with the spread. An important observation in the spread tables (Table 2,

Table 3, and Table 4) is that the model risk estimates of the parameters do not sum up to the 'global' model risk. In other words, the total model risk is not the sum of the parameter contributions to the model risk. This can best be understood with a visual example shown in Figure 5.

We see that the spread measures for estimates with fixed parameter values are less than or equal to the 'global' (referred to as 'all' in the figure) spread measure. Since each point refers to a parameter combination of four parameters, there is an overlap in their distributions, and the spread metric for individual parameters cannot sum up to the overall spread. This implies that the parameter contributions are not marginal contributions, and it makes sense only to look at parameter contributions among each other for comparison, and not as true contributions to the observed variation. This is a limitation of the spread metric, and it would be useful to devise model risk metrics that can satisfy this property to create a 'true marginal parameter contribution'. However, within the scope of this work, the spread metric is already useful for ranking parameter contributions, summarized in Table 5.

The parameter contributing the most to model risk is usually the historical period parameter (except for the ΔCoVaR at 99% confidence level). This parameter is linked to the memory of past events and has the most variation linked to how quickly large price jumps are forgotten. The parameter contributing the least is the window type parameter (except for the MES at 95% confidence level), indicating that the shape of the historical window used to determine statistics is less important. The distribution type parameter contributes more to the model risk than the inter-FI dependencies parameter, highlighting the higher importance of the modelling of the tail behaviour over the type of correlation matrix. This type of analysis can also be done for the four DGP-models, in which case, we have only two parameters to rank. From Table 3 and Table 4, we see that the historical period parameter contributes more than the window type for both SRMs under both risk conditions (95% and 99% confidence levels).

The approach to rank parameter contributions presented here is general enough that it can be extended to any parametric algorithm. This can be a useful tool for practitioners to analyse the algorithms themselves and potential impact of various parameter choices on the model results.

In addition to estimating the model risk per bank, we looked at how the assignment of the banks to risk buckets might alter with changes to the model parameters. As explained in the previous section (Section 2.5), we performed a ranking of each FI for every parameter combination and assigned them to risk buckets B0–B5, with B0 consisting of banks having the lowest risk (1/6$^{\text{th}}$ smallest absolute values of the considered SRM), B5 consisting of banks having the highest risk (1/6$^{\text{th}}$ largest absolute values).

Figure 6 contains an example of 30 May 2018 of the bucketing for two banks (Société Générale, and Raiffeisen Bank International) based on their MES (left) and ΔCoVaR (middle) at 95% confidence level across all 24 model estimates (right subplot contains explanation of the parameter combinations). We observed that Société Générale can be assigned anywhere from B1 to B5 based on the MES, and anywhere from B0 to B5 based on the ΔCoVaR. Raiffeisen Bank International gets assigned anywhere from B1 to B5 based on its MES, and anywhere from B0 to B4 based on its ΔCoVaR. To verify whether this level of disagreement comes because of comparing different distributions, we considered the same example in Figure 7, but focus on a single DGP: multivariate Student's-t. The number of parameter combinations is now reduced from 24 to 6, but we still see Société Générale being assigned anywhere from B1 to B4 (B2 to B5) based its MES (ΔCoVaR), and Raiffeisen Bank International being assigned anywhere from B1 to B5 (B0 to B4) based on its MES (ΔCoVaR). This case highlights that even with just six parameter combinations and six risk buckets, we have banks that might be classified as low risk (B0/B1) with one parameter setting, and high risk (B4/B5) with another parameter setting. In both cases, we see that there is quite a lot of disagreement among the parameter combinations, indicating that the observed model risk for a single FI does propagate across to the ranking and the clustering of the banks.

To quantify the bucketing concordance, we computed the daily Krippendorff's-alpha and the percentage agreement across the 24 parameter combinations (no DGP-split analysis was done here). The monthly averages (for visibility) of the agreement metrics are plot in Figure 8. The Krippendorff's--alpha (top subplot) and percentage agreement (bottom subplot) indicate that the agreement for the MES is better than that for the ΔCoVaR. As with the spread, we observed that both metrics were lower for the higher risk setting (99% confidence level) than for the lower risk setting (95% confidence level). The summary statistics of the two metrics are provided in Table 6. The Krippendorff's-alpha metric in our case has a notion of distances between buckets, i.e. B0/B5 has a 'larger' distance than B0/B2, which can explain the perceived higher values when compared to the percentage agreement metric, which treats all mismatches in the same way.

Our results highlight the limitations of these SRMs, especially important for regulatory purposes. High dispersion values of the SRMs indicate a low reliability, making it difficult to truly judge the level of stress within the financial network. The risk is not completely removed after ranking or bucketing, since banks can be classified as extremely risky, or completely safe by changing a few parameters. Regulatory frameworks require banks to keep an extra capital buffer based on the assigned risk bucket, and parameterization error alone can misclassify a bank. A high-risk bank misplaced in a lower bucket will increase the level of systemic stress within the financial system and in the event of a systemic trigger, the bank might not have a sufficient buffer to protect itself and its dependencies from the shock. On the other hand, a low-risk bank misplaced in a higher bucket will penalize the bank by requiring an unnecessary capital buffer and stop it from taking on potentially interesting but risky ventures. While this might not seem as dangerous as the first, banks are required for economic stimulus and growth, and inhibiting banks too much might cause the modern economy to slow down, creating systemic stresses elsewhere.

4. Conclusions and perspectives

Systemic risk plays a significant role in modern finance and is a useful tool for regulators. The model risk of SR is equally vital, since it measures the risk associated with an inaccurate reporting of the SR measures. In this work, we addressed the problem of quantifying the model risk of two SRMs, MES and ΔCoVaR, by using the spread as a model risk metric, for a set of FIs deemed systemically important by the ECB. The model risk source specifically considered was the parametrization error of the MC process used to compute SRM estimates. The spread also allows one to rank the parameter contributions to the observed model risk by computing drops in model risk of estimates with fixed parameter values. Larger average drops over parameter values implied higher contributions. This approach is easily extendible to other parametric algorithms and can help analyse sensitivities of a model to different parameters.

To estimate the model risk, multiple estimates for each of the SR measures were generated by altering the parameters in the MC algorithm. The model risk computed with this scheme is very closely related to quantifying the estimation risk and is a complementary approach to a data-driven method that relies on bootstrapping data without modifying the model parameters. Our proposed approach builds on the state of the art by being computationally cheaper than the data driven approach, since it relies on a smaller number of samples (24 in our case) than sampling data (typically in the hundreds or thousands).

Our observations can be summarized as follows. Firstly, the spread estimates were always above zero, and right skewed, implying that model risk is always present and can be exceptionally large. Secondly, model risk increases with an increase in risk condition, which is intuitive and can be interpreted as being due to at least two factors: (i) higher confidence levels focus more on the tails, and hence depend more on the modelling of the tail behaviour, and (ii) fewer MC iterations satisfy the stricter conditions needed for higher confidence levels, leading to an increased variability of the SRM values. Thirdly, model risk typically increases after volatile periods, because the SRM estimates react to the drop in volatilities at different speeds. Fourthly, ΔCoVaR has a higher model risk than MES, indicating an increase of model risk with computational complexity. We ranked the parameter contributions to find that the parameter that contributed most was the historical period, implying that the memory (longer windows take more time to forget extreme events) and reactivity (shorter windows react faster to latest information) are the main factors to explain the observed model risk. The parameter contributing the least was the window type parameter, implying that the actual shape (weights to individual data points) of the window mattered less. Finally, we provided a bucketing of the banks into six risk buckets and showed that the model risk is not completely removed, and that misclassification of the bank becomes quite important.

There are multiple implications associated with our findings. The model risk tends to increase when the risk condition is stricter. This is important, notably in the context of banking regulation. Systemic events have more real-life consequences than non-systemic events. A larger model risk for SR indicates a lower reliability of the measure, thereby making it difficult for regulators to really judge the real amount of stress within the financial system. This difficulty is further compounded by an increase in model risk with an increase in risk condition. A stricter risk condition tries to evaluate the reaction of the FIs and the market to events that are more extreme. The larger model risk implies a reduced capability for regulators to judge the actual SR, since the systemically important FIs (SIFIs) as deduced from a model may not be the same when the parameters are slightly altered. Misidentification of SIFIs is of critical concern to FIs and regulators, as: (i) underestimating the risk can increase the level of stress within the system, whereas (ii) overestimating the risk can decrease the investment stimulus from FIs, both of which are not ideal for the economy.

These arguments allow us to propose some simple suggestions that use the omnipresent model risk. A straight-forward way is to report average stress test results along with the model risk (or to report the multiple estimates directly), so that the distribution of the estimates can be analysed to identify SIFIs more accurately. A more in-depth analysis of how estimation risk affects the level of stress within a financial network can be highly informative. For example, if the MES of an FI is reported along with the model risk, then one can compute the MES on pessimistic and optimistic scenarios, where the MES is increased and decreased respectively by the model risk amount. By collecting the MES estimates on three scenarios (pessimistic, average, optimistic), it might be possible to generate a 'systemic stress distribution' based on different scenarios at individual FIs. By additionally combining these distributions of multiple SRMs (e.g. MES and ΔCoVaR), multiple SR effects can be studied, which can be more instructive to maintain the stability of financial systems.

References

Acharya V.V., Pedersen L.H., Philippon T., Richardson M.P. (2011), *Measuring systemic risk*, AFA 2011 Denver Meetings Paper.

Adrian T., Brunnermeier M.K. (2008), *CoVaR*, Staff Reports, 348, Federal Reserve Bank of New York, https://ideas.repec.org/p/fip/fednsr/348.html.

Benoit S., Colliard J.-E., Hurlin C., Pérignon C. (2017), Where the risks lie: a survey on systemic risk, *Review of Finance*, 21(1), 109–152.

Bianchi M.L., Sorrentino A.M. (2020), Measuring CoVaR: an empirical comparison, *Computational Economics*, 55, 511–528.

Boucher C.M., Danielsson J., Kouontchou P.S., Maillet B.B. (2014), Risk models-at-risk, *Journal of Banking & Finance*, 44, 72–92.

Brownlees C.T., Engle R. (2012), *Volatility, correlation and tails for systemic risk measurement,* available at SSRN, 1611229.

Castro C., Ferrari S. (2014), Measuring and testing for the systemically important financial institutions, *Journal of Empirical Finance*, 25, 1–14.

Castro S. (2017), GitHub, Fast Krippendorff: Fast computation of Krippendorff's alpha agreement measure, https://github.com/pln-fing-udelar/fast-krippendorff, retrieved December 2024.

Christoffersen P., Gonçalves S. (2004), *Estimation risk in financial risk management*, CIRANO Working Papers, 2004s-15, https://EconPapers.repec.org/RePEc:cir:cirwor:2004s-15.

Crouhy M., Galai D., Mark R. (1998), Model risk, *Journal of Financial Engineering*, 7, 267–288.

Danielsson J., James K.R., Valenzuela M., Zer I. (2016a), Can we prove a bank guilty of creating systemic risk? A minority report, *Journal of Money, Credit and Banking*, 48, 795–812.

Danielsson J., James K.R., Valenzuela M., Zer I. (2016b), Model risk of risk models, *Journal of Financial Stability,* 23, 79–91.

Derman E. (1996), *Model risk*, Quantitative Strategies Research Notes, Goldman Sachs.

EuSpRIG (2013), *EuSPRIG Horror Stories*, European Spreadsheet Risk Interest Group, https://eusprig.org/research-info/horror-stories/.

Gelman A., Carlin J.B., Stern H.S., Rubin D.B. (1995), *Bayesian Data Analysis*, Chapman and Hall/CRC.

Glasserman P., Xu X. (2014), Robust risk measurement and model risk, *Quantitative Finance*, 14, 29–58.

Green P.J. (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, 82(4), 711.

Hendricks D. (1996), Evaluation of value-at-risk models using historical data, *Economic Policy Review*, 2, 39–69.

Hoeting J.A., Madigan D., Raftery A.E., Volinsky C.T. (1999), Bayesian model averaging: a tutorial, *Statistical Science*, 14(4), 382–417.

Holton G.A. (2013), *Value-at-Risk. Theory and Practice*, https://www.value-at-risk.net/.

Hurd T.R. (2016), *Contagion! Systemic Risk in Financial Networks*, Springer.

Idier J., Lamé G., Mésonnier J.-S. (2013), *How useful is the marginal expected shortfall for the measurement of systemic exposure? A practical assessment*, ECB Working Paper, 1546, European Central Bank.

Jajuga K. (2014), Systemic risk and financial stability in insurance: macroprudential policy concerns, in: J. Monkiewicz, M. Małecki (eds.), *Macroprudential Supervision in Insurance: Theoretical and Practical Aspects*, Palgrave Macmillan.

Jorion P. (1996), Risk2: measuring the risk in Value at Risk, *Financial Analysts Journal*, 52, 47–56, DOI: 10.2469/faj.v52.n6.2039.

Kaufman G.G. (1995), Comment on systemic risk, in: G. Kaufman (ed.), *Research in Financial Services: Banking, Financial Markets, and Systemic Risk*, vol. 7, JAI Press.

Kaufman G.G., Scott K.E. (2003), What is systemic risk, and do bank regulators retard or contribute to it?, *The Independent Review*, 7, 371–391.

Klein R.W., Bawa V. S. (1976), The effect of estimation risk on optimal portfolio choice, *Journal of Financial Economics*, 3, 215–231, DOI: 10.1016/0304-405X(76)90004-0.

Krippendorff K. (2019), *Content Analysis: An Introduction to Its Methodology*, SAGE Publications.

Lewellen J., Shanken J. (2000), *Estimation risk, market efficiency, and the predictability of returns*, NBER Working Paper, 7699, National Bureau of Economic Research.

Löffler G., Raupach P. (2018), Pitfalls in the use of systemic risk measures, *Journal of Financial and Quantative Analysis*, 53(1), 269–298.

Mishkin F. (1995), Comment on systemic risk, in: G. Kaufman (ed.), *Research in Financial Services: Banking, Financial Markets*, *and Systemic Risk*, 7, 31–45, JAI Press.

Pasieczna A.H. (2019), Monte Carlo simulation approach to calculate Value at Risk: application to WIG20 and MWIG40, *Financial Sciences. Nauki o Finansach*, 24(2), 61–75.

Pasieczna A.H. (2021), Model risk of VaR and ES using Monte Carlo: study on financial institutions from Paris and Frankfurt stock exchanges, in: K. Jajuga, H. Locarek-Junge, L.T. Orlowski, K. Staehr (eds.), *Contemporary Trends and Challenges in Finance*, Springer.

Pasieczna A.H., Szydłowska M.J. (2021), Estimating model risk of VaR under different approaches: study on European banks, *Współczesne Problemy Zarządzania*, 9, 65–76, DOI: 10.52934/wpz.151.

Pollack L. (2013), A tempest in a spreadsheet, *Financial Times*, https://www.ft.com/content/80bc5046-6ed2-30ef-b31a-bcbeac55d8bb.

Robert C.P., Casella G. (2013), *Monte Carlo Statistical Methods*, Springer.

Roncoroni A., Battiston S., D'Errico M., Hałaj G., Kok C. (2019), *Interconnected banks and systemically important exposures*, SSRN, 3491235.

Schwarcz S.L. (2008), Systemic risk, *Georgetown Law Journal*, 97(1), 193.

Sibbertsen P., Stahl G., Luedtke C. (2008), Measuring model risk, *Journal of Risk Model Validation*, 2, 65–82.

Virtanen P. et al. (2020), SciPy 1.0: fundamental algorithms for scientific computing in Python, *Nature Methods*, 17, 261–272, DOI: 10.1038/s41592-019-0686-2.

Wagner W., Fisher E., Pascual P. (2010), Misunderstanding models in environmental and public health regulation, *New York University Environmental Law Journal*, 18, 293.

Acknowledgements

I would like to thank Kozminski University for access to the Bloomberg Terminal.

Appendix

Table 1
List of FIs in the study

Country code	Name	Grounds of significance
BE	AXA Bank Belgium SA; AXA Bank Belgium NV	Article 6(5)(b) of Regulation (EU) No. 1024/2013
	KBC Group NV	Size (total assets EUR 150–300 bn)
BG	DSK Bank AD	Among the three largest credit institutions in the Member State
DE	Aareal Bank AG	Size (total assets EUR 30–50 bn)
	Commerzbank AG	Size (total assets EUR 300–500 bn)
	Deutsche Bank AG	Size (total assets above EUR 1,000 bn)
	Deutsche Pfandbriefbank AG	Size (total assets EUR 50–75 bn)
EE	AS SEB Pank	Total assets above 20% of GDP
	Swedbank AS	Total assets above 20% of GDP
IE	AIB Group plc	Size (total assets EUR 75–100 bn)
	Bank of Ireland Group plc	Size (total assets EUR 100–150 bn)
GR	Alpha Services and Holdings S.A.	Size (total assets EUR 50–75 bn)
	Eurobank Ergasias Services and Holdings S.A.	Size (total assets EUR 50–75 bn)
	National Bank of Greece S.A.	Size (total assets EUR 50–75 bn)
	Piraeus Financial Holdings S.A.	Size (total assets EUR 50–75 bn)
ES	Banco Bilbao Vizcaya Argentaria S.A.	Size (total assets EUR 500–1,000 bn)
	Banco de Sabadell S.A.	Size (total assets EUR 150–300 bn)
	Banco Santander S.A.	Size (total assets above EUR 1,000 bn)
	Bankinter S.A.	Size (total assets EUR 75–100 bn)
	CaixaBank S.A.	Size (total assets EUR 300–500 bn)
	Unicaja Banco S.A.	Size (total assets EUR 50–75 bn)
FR	BNP Paribas S.A.	Size (total assets above EUR 1,000 bn)
	Crédit Agricole S.A.	Size (total assets above EUR 1,000 bn)
	Société Générale S.A.	Size (total assets above EUR 1,000 bn)
IT	Banca Carige S.p.A. – Cassa di Risparmio di Genova e Imperia	Article 6(5)(b) of Regulation (EU) No. 1024/2013
	Banca Monte Dei Paschi di Siena S.p.A.	Size (total assets EUR 100–150 bn)
	Banca Popolare di Sondrio, Società Cooperativa per Azioni	Size (total assets EUR 30–50 bn)
	Banco BPM S.p.A.	Size (total assets EUR 150–300 bn)
	BPER Banca S.p.A.	Size (total assets EUR 75–100 bn)
	Intesa Sanpaolo S.p.A.	Size (total assets EUR 500–1,000 bn)

Table 1, cont'd

	Mediobanca – Banca di Credito Finanziario S.p.A.	Size (total assets EUR 75–100 bn)
	UniCredit S.p.A.	Size (total assets EUR 500–1,000 bn)
CY	Bank of Cyprus Holdings PLC	Total assets above 20% of GDP
	Hellenic Bank PLC	Total assets above 20% of GDP
LT	Akcinė bendrovė Šiaulių bankas	Among the three largest credit institutions in the Member State
MT	Bank of Valletta p.l.c.	Total assets above 20% of GDP
	HSBC Bank Malta p.l.c.	Total assets above 20% of GDP
NL	ABN AMRO Bank N.V.	Size (total assets EUR 300–500 bn)
	ING Groep N.V.	Size (total assets EUR 500–1,000 bn)
AT	Addiko Bank AG	Significant cross-border activities
	BAWAG Group AG	Size (total assets EUR 30–50 bn)
	Erste Group Bank AG	Size (total assets EUR 150–300 bn)
	Raiffeisen Bank International AG	Size (total assets EUR 150–300 bn)
	Sberbank Europe AG	Significant cross-border activities
PT	Banco Comercial Português, SA	Size (total assets EUR 75–100 bn)
SI	Nova Ljubljanska Banka d.d. Ljubljana	Total assets above 20% of GDP
FI	Nordea Bank Abp	Size (total assets EUR 500–1,000 bn)

Table 2

Spread values for the MES and ΔCoVaR at 95% and 99% over all estimates and estimates with one parameter *P* fixed

SRM	Fixed value	95%	$\Delta_P^{95\%}$	$\bar{\Delta}_P^{95\%}$	99%	$\Delta_P^{99\%}$	$\bar{\Delta}_P^{99\%}$
MES	All	0.677			1.137		
	Inter-FI dependencies						
	independent	0.660	-0.017	-0.096	1.113	-0.024	-0.255
	multivariate	0.503	-0.174		0.651	-0.486	
	Distribution type						
	Student's-t	0.593	-0.084	-0.189	1.025	-0.112	-0.373
	normal	0.384	-0.293		0.504	-0.633	
	Window type						
	rolling	0.587	-0.090	-0.136	1.002	-0.135	-0.190
	EWM	0.495	-0.182		0.893	-0.244	
	Historical period						
	125D	0.336	-0.341		0.592	-0.545	
	250D	0.391	-0.286	-0.289	0.710	-0.427	-0.432
	500D	0.437	-0.240		0.814	-0.323	
ΔCoVaR	All	0.854			1.850		
	Inter-FI dependencies						
	independent	0.805	-0.049	-0.154	1.794	-0.056	-0.519
	multivariate	0.594	-0.260		0.868	-0.982	
	Distribution type						
	Student's-t	0.734	-0.120	-0.267	1.687	-0.163	-0.678
	normal	0.440	-0.414		0.657	-1.193	
	Window type						
	rolling	0.733	-0.121	-0.139	1.617	-0.233	-0.240
	EWM	0.697	-0.157		1.603	-0.247	
	Historical period						
	125D	0.469	-0.385		1.033	-0.817	
	250D	0.564	-0.290	-0.294	1.290	-0.560	-0.566
	500D	0.647	-0.207		1.530	-0.320	

$\Delta_P^{x\%}$ denotes the drop in spread and $\bar{\cdot}$ is the average.

Table 3
Spread values for the MES at 95% and 99% separated by the data-generating processes, over all estimates and estimates with one parameter P fixed

DGP	Fixed value	95%	$\Delta_P^{95\%}$	$\overline{\Delta}_P^{95\%}$	99%	$\Delta_P^{99\%}$	$\overline{\Delta}_P^{99\%}$
Normal distribution, independent FIs	All	0.354			0.455		
	Window type						
	rolling	0.283	-0.071	-0.115	0.360	-0.095	0.148
	EWM	0.194	-0.160		0.254	-0.201	
	Historical period						
	125D	0.115	-0.239		0.149	-0.306	
	250D	0.126	-0.228	-0.231	0.165	-0.290	-0.294
	500D	0.129	-0.225		0.168	-0.287	
Normal multivariate distribution	All	0.359			0.460		
	Window type						
	rolling	0.286	-0.073	-0.118	0.363	-0.097	-0.150
	EWM	0.197	-0.162		0.257	-0.203	
	Historical period						
	125D	0.116	-0.243		0.149	-0.311	
	250D	0.128	-0.231	-0.234	0.166	-0.294	-0.298
	500D	0.131	-0.228		0.172	-0.288	
Student's-t distribution, independent FIs	All	0.501			0.791		
	Window type						
	rolling	0.403	-0.098	-0.146	0.639	-0.152	-0.214
	EWM	0.307	-0.194		0.515	-0.276	
	Historical period						
	125D	0.138	-0.363		0.199	-0.592	
	250D	0.159	-0.342	-0.346	0.234	-0.557	-0.563
	500D	0.168	-0.333		0.251	-0.540	
Student's-t multivariate distribution	All	0.422			0.549		
	Window type						
	rolling	0.337	-0.085	-0.133	0.434	-0.115	-0.173
	EWM	0.240	-0.182		0.319	-0.230	
	Historical period						
	125D	0.128	-0.294		0.169	-0.380	
	250D	0.145	-0.277	-0.281	0.193	-0.356	-0.361
	500D	0.151	-0.271		0.202	-0.347	

$\Delta_P^{x\%}$ denotes the drop in spread and $\overline{\cdot}$ is the average.

Table 4

Spread values for the ΔCoVaR at 95% and 99% separated by the data-generating processes, over all estimates and estimates with one parameter P fixed

DGP	Fixed value	95%	$\Delta_P^{95\%}$	$\mathbf{\Delta_P^{95\%}}$	99%	$\Delta_P^{99\%}$	$\mathbf{\Delta_P^{99\%}}$
Normal distribution, independent FIs	All	0.360			0.534		
	Window type						
	rolling	0.269	-0.091	-0.117	0.394	-0.140	-0.174
	EWM	0.217	-0.143		0.326	-0.208	
	Historical period						
	125D	0.126	-0.234		0.191	-0.343	
	250D	0.143	-0.217	-0.221	0.214	-0.320	-0.325
	500D	0.148	-0.212		0.221	-0.313	
Normal multivariate distribution	All	0.372			0.547		
	Window type						
	rolling	0.280	-0.092	-0.120	0.407	-0.140	-0.177
	EWM	0.223	-0.149		0.333	-0.214	
	Historical period						
	125D	0.129	-0.243		0.194	-0.353	
	250D	0.146	-0.226	-0.230	0.217	-0.330	-0.335
	500D	0.152	-0.220		0.226	-0.321	
Student's-t distribution, independent FIs	All	0.559			1.155		
	Window type						
	rolling	0.426	-0.133	-0.159	0.891	-0.264	-0.294
	EWM	0.375	-0.184		0.832	-0.323	
	Historical period						
	125D	0.155	-0.404		0.276	-0.879	
	250D	0.185	-0.374	-0.378	0.337	-0.818	-0.824
	500D	0.203	-0.356		0.380	-0.775	
Student's-t multivariate distribution	All	0.485			0.725		
	Window type						
	rolling	0.367	-0.118	-0.152	0.538	-0.187	-0.232
	EWM	0.298	-0.187		0.448	-0.277	
	Historical period						
	125D	0.156	-0.329		0.242	-0.483	
	250D	0.182	-0.303	-0.308	0.281	-0.444	-0.451
	500D	0.194	-0.291		0.300	-0.425	

$\Delta_P^{x\%}$ denotes the drop in spread and · is the average.

Table 5

Ranking of parameter contributions to model risk

Ranking	MES		ΔCoVaR	
	95%	99%	95%	99%
1	HP	HP	HP	DT
2	DT	DT	DT	HP
3	WT	IFD	IFD	IFD
4	IFD	WT	WT	WT

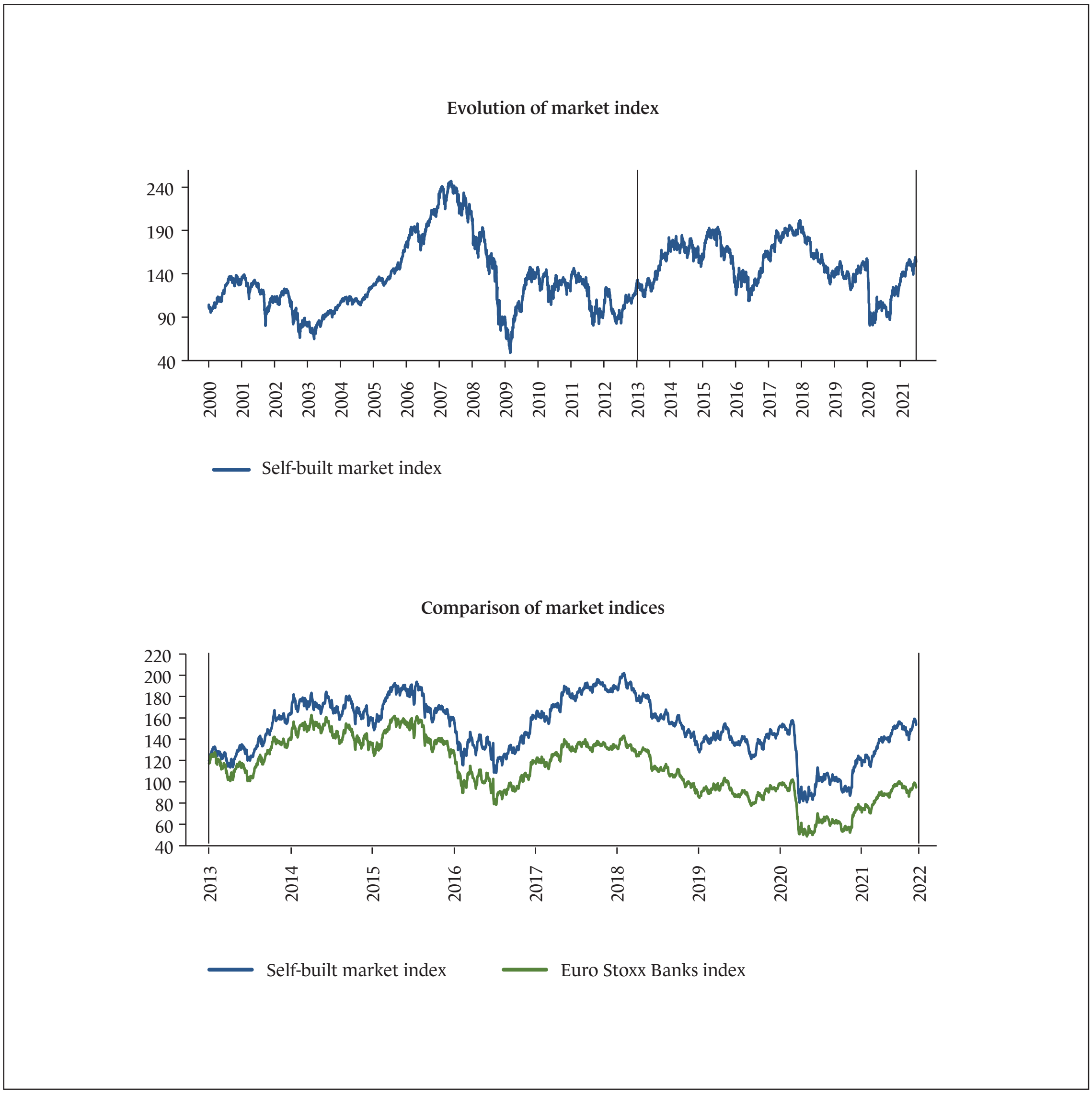
Notes:
The parameter contributing the most to the model risk is ranked as 1 and that contributing the least is ranked as 4. Abbreviations: historical period (HP), distribution type (DT), window type (WT), inter-FI dependencies (IFD).

Table 6

Statistics of the agreement metrics (Krippendorff's-alpha and percentage agreement) for the MES and ΔCoVaR at 95% and 99% confidence levels

SRM		Krippendorff's-alpha			Percentage agreement		
		mean (std)	minimum	maximum	mean (std)	minimum	maximum
MES	95%	0.86 (0.08)	0.60	0.98	0.59 (0.11)	0.37	0.89
	99%	0.85 (0.08)	0.59	0.98	0.57 (0.10)	0.37	0.86
ΔCoVaR	95%	0.43 (0.08)	0.09	0.74	0.28 (0.03)	0.20	0.41
	99%	0.28 (0.07)	0.03	0.65	0.25 (0.02)	0.18	0.39

Figure 1
Evolution of the market index built from the FIs within the dataset and comparison with the Euro Stoxx Banks index



Note: the starting value of the self-built index is 100 points, and index values represent the growth with respect to this value.

Figure 2
Distribution of relative market capitalization values of all FIs (last simulation day) and distribution of the inter-FI price return correlation values (last simulation year)

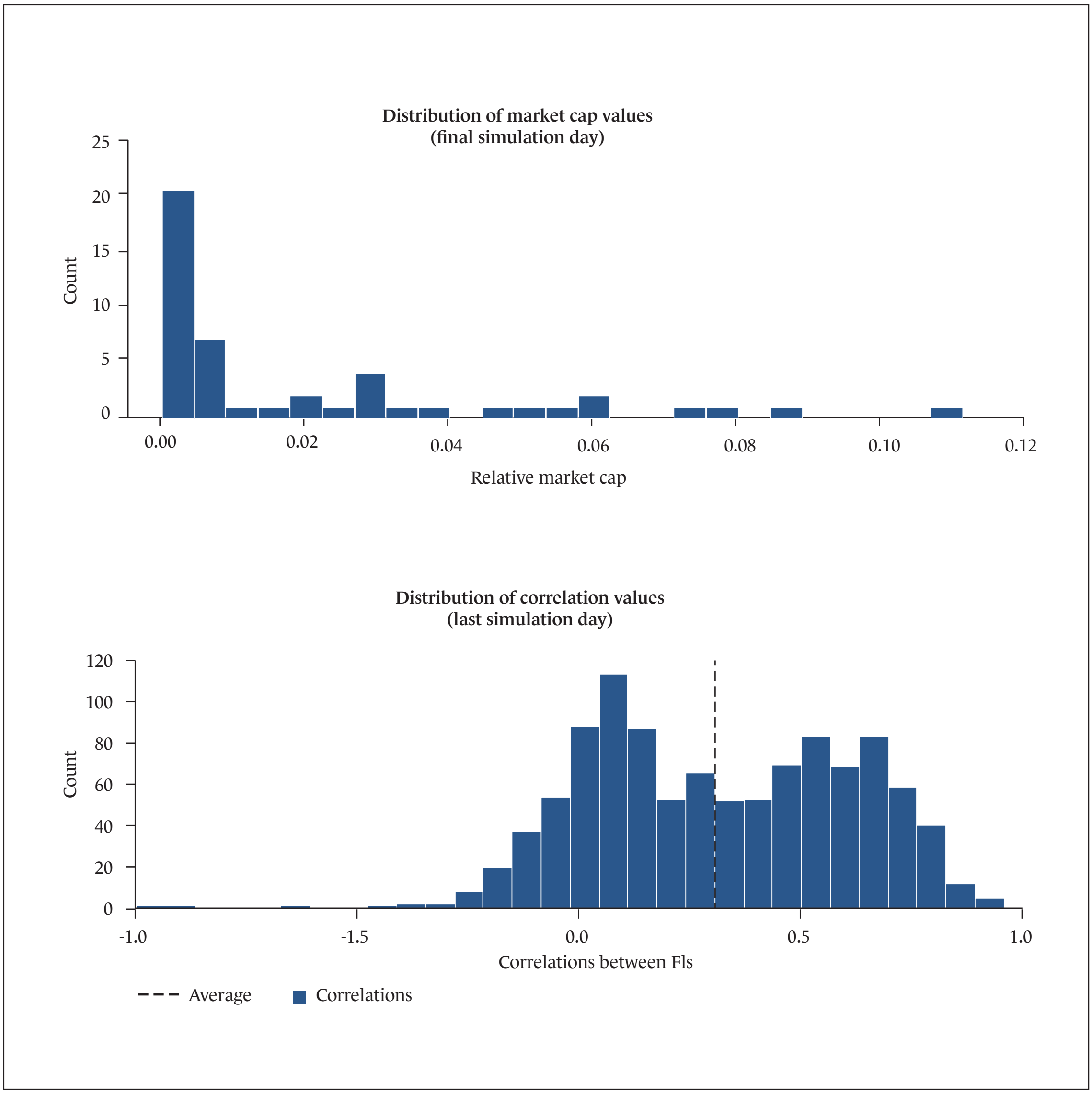


Figure 3
Model risk of MES at 95% and 99% confidence levels for all FIs

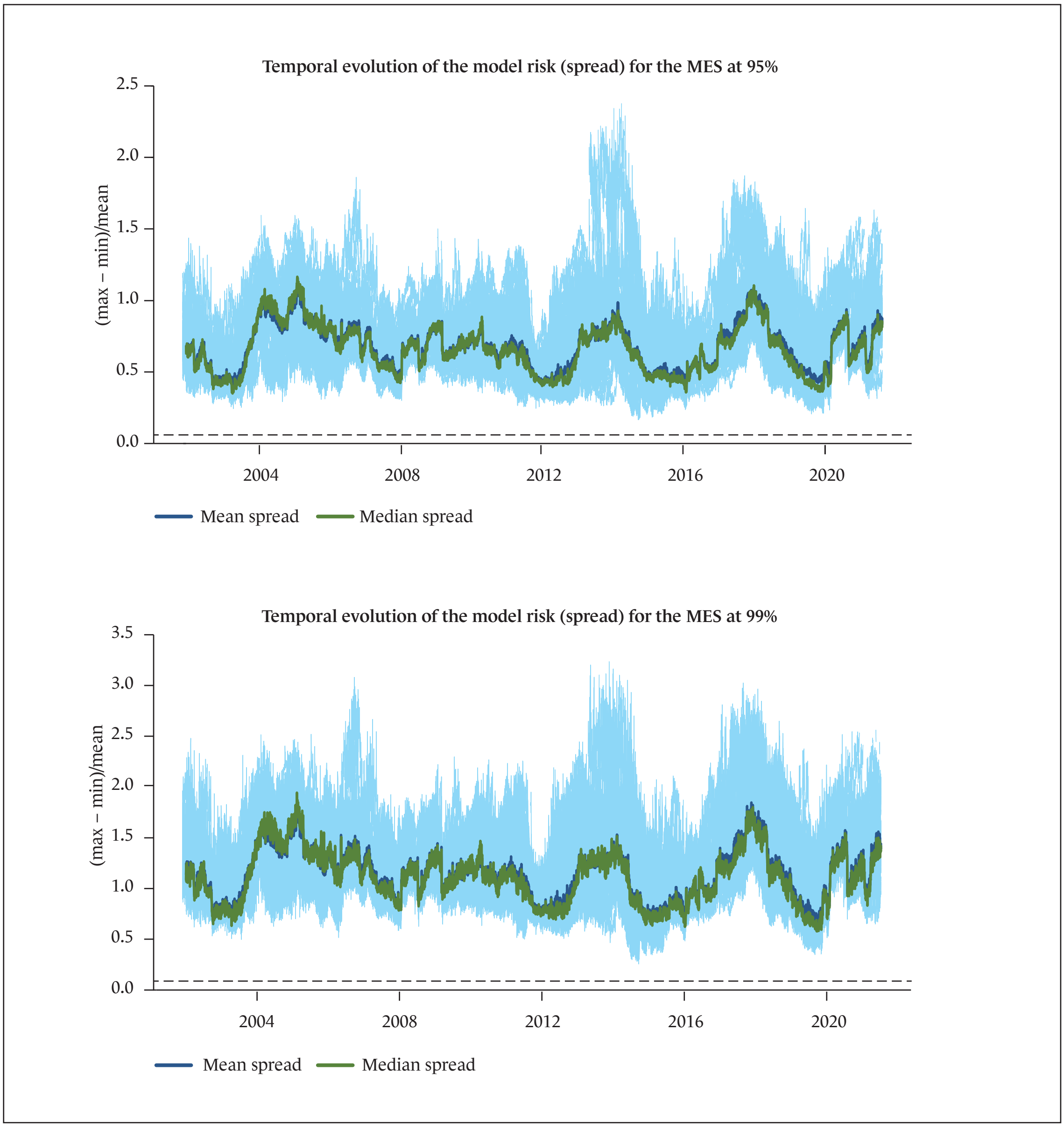


Figure 4
Model risk of ΔCoVaR at 95% and 99% confidence levels for all FIs

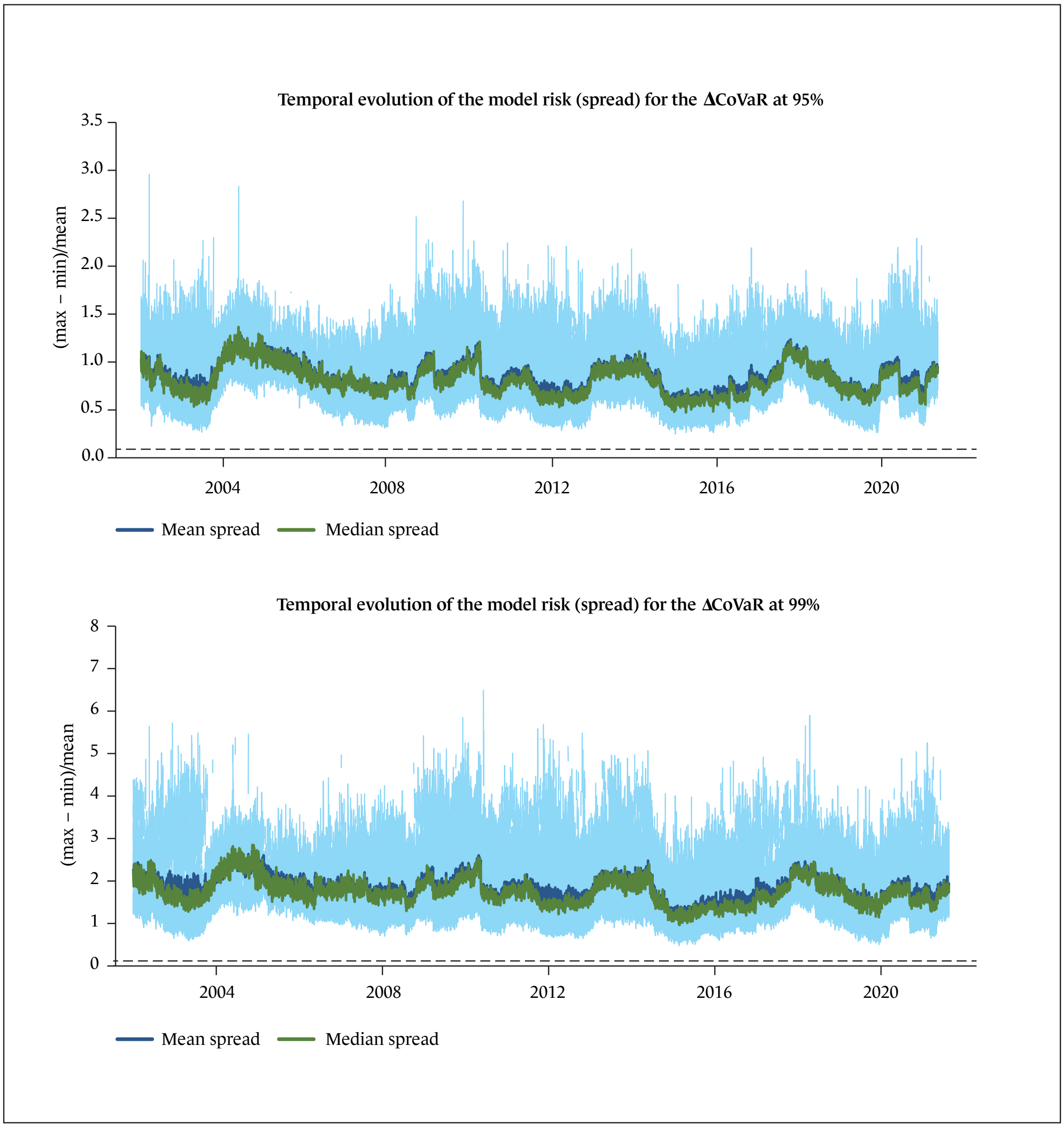
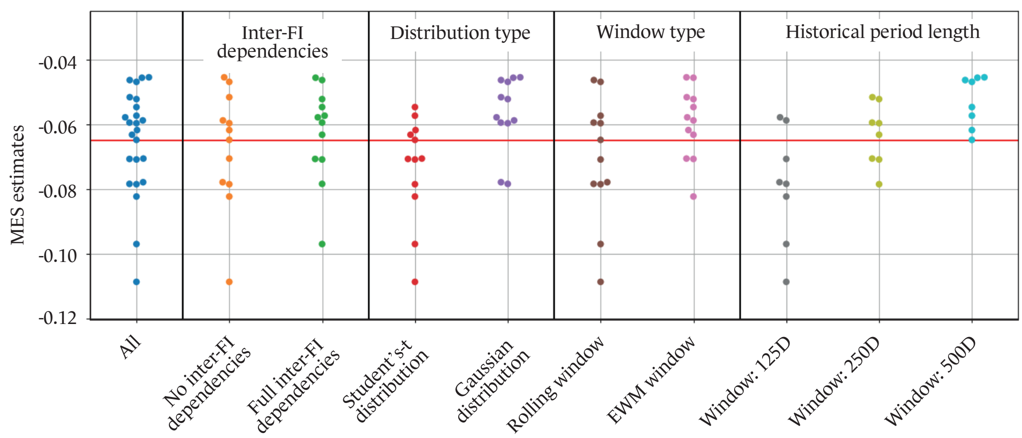
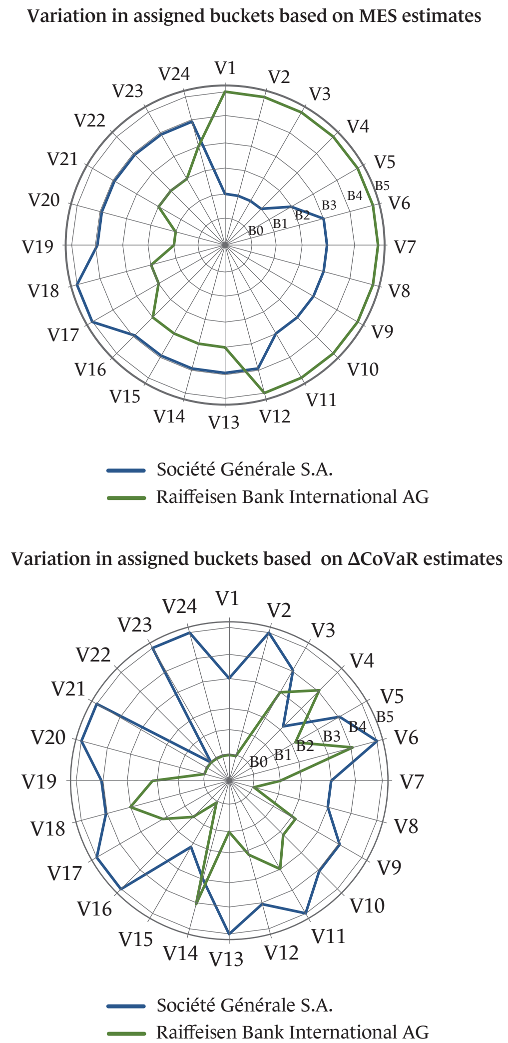


Figure 5
Example of a distribution of 24 MES estimates as percentage returns



Notes:
Losses are negative.
BNP Paribas was the FI chosen, and the predictions were for 27 August 2020.

Figure 6
Example of the assignment of two banks (Société Générale and Raiffeisen Bank International) for MES at 95%, and ΔCoVaR at 95% across all 24 model estimates



Variant information

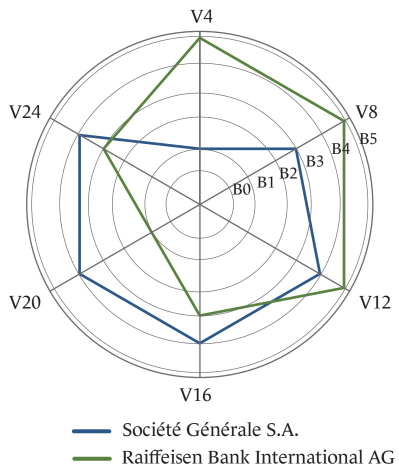
V1	Normal/Univariate: rolling/125D
V2	Student's-t/Univariate: rolling/125D
V3	Normal/Multivariate: rolling/125D
V4	Student's-t/Multivariate: rolling/125D
V5	Normal/Univariate: ewm/125D
V6	Student's-t/Univariate: ewm/125D
V7	Normal/Multivariate: ewm/125D
V8	Student's-t/Multivariate: ewm/125D
V9	Normal/Univariate: rolling/250D
V10	Student's-t/Univariate: rolling/250D
V11	Normal/Multivariate: rolling/250D
V12	Student's-t/Multivariate: rolling/250D
V13	Normal/Univariate: ewm/250D
V14	Student's-t/Univariate: ewm/250D
V15	Normal/Multivariate: ewm/250D
V16	Student's-t/Multivariate: ewm/250D
V17	Normal/Univariate: rolling/500D
V18	Student's-t/Univariate: rolling/500D
V19	Normal/Multivariate: rolling/500D
V20	Student's-t/Multivariate: rolling/500D
V21	Normal/Univariate: ewm/500D
V22	Student's-t/Univariate: ewm/500D
V23	Normal/Multivariate: ewm/500D
V24	Student's-t/Multivariate: ewm/500D

Note: B5 corresponds to the bucket with the highest risk. The predictions were for 30 May 2018.

Figure 7

Example of the assignment of two banks (Société Générale and Raiffeisen Bank International) for MES at 95%, and ΔCoVaR at 95% across the six model estimates for the Student's-t multivariate distribution

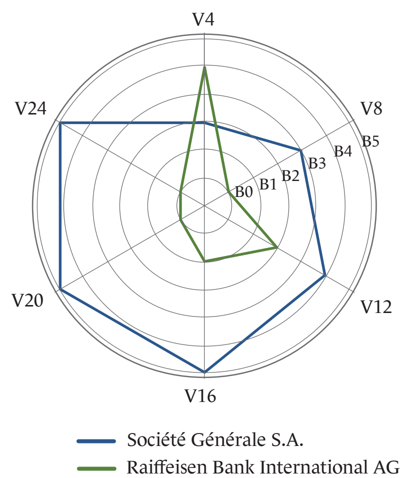
Variation in assigned buckets based on MES estimates



Variant information

V4	Student's-t/Multivariate: rolling/125D
V8	Student's-t/Multivariate: ewm/125D
V12	Student's-t/Multivariate: rolling/250D
V16	Student's-t/Multivariate: ewm/250D
V20	Student's-t/Multivariate: rolling/500D
V24	Student's-t/Multivariate: ewm/500D

Variation in assigned buckets based on ΔCoVaR estimates



Note: B5 corresponds to the bucket with the highest risk. The predictions were for 30 May 2018.

Figure 8
Monthly average values of Krippendorff's alpha and percentage agreement of the FI bucketing for MES and ΔCoVaR at 95% and 99%

Note: value of 1 indicates complete agreement, and value of 0 indicates complete disagreement.

Jak wiarygodne są miary ryzyka systemowego? Oszacowanie ryzyka modelu MES i ΔCoVaR

Streszczenie

Miary ryzyka systemowego (MRS) są kluczowe w analizie stresu w systemie finansowym. Niedoszacowanie ryzyka systemowego zwiększa stres w sieci, podczas gdy przeszacowanie może prowadzić do nakładania na banki nadmiernych kar. Na wiarygodność tych miar ryzyka systemowego wpływa jednak ryzyko modelu, definiowane w niniejszej pracy jako niepewność wyników modelu wynikająca z tzw. błędu parametryzacji. Wcześniejsze badania (Danielsson i in. 2016b) zwracały uwagę na znaczną zmienność wyników MRS, przy czym niektóre z nich wykazywały niewiarygodne rankingi instytucji finansowych. W naszej pracy oparliśmy się na tych ustaleniach, kwantyfikując ryzyko modelowe dla dwóch MRS, Marginal Expected Shortfall (MES) i Delta Conditional Value at Risk (ΔCoVaR), na przykładzie systemowo ważnych banków europejskich. Ponadto pokazaliśmy wpływ tego modelu na rozbieżność między rankingami znaczenia systemowego banków. Celem pracy jest lepsze zrozumienie ryzyka modelowego w aspekcie MRS oraz tego, jak wybór parametrów może wpływać na ranking instytucji finansowych i interpretację ryzyka systemowego.

Badanie ma na celu omówienie ryzyka modelu w MRS i jego konsekwencji dla oceny ryzyka systemowego. W szczególności w badaniu postawiono pytania: w jaki sposób zmienność parametrów modelu oddziałuje na oszacowania MES i ΔCoVaR dla systemowo ważnych banków europejskich oraz jak ryzyko modelu oddziałuje na rankingi ryzyka systemowego banków. Dodatkowo zbadano, jak można kwantyfikować ryzyko modelu i klasyfikować je pod względem wrażliwości parametrów, aby zapewnić większą wiarygodność decyzji opartych na MRS.

Ryzyko modelu istnieje zarówno w MES, jak i w ΔCoVaR ze względu na wrażliwość tych miar na zmiany parametrów w ich szacunkach, w tym przypadku przy użyciu techniki Monte Carlo (MC). Wpływa ono na ranking ryzyka systemowego banków. Ryzyko modelu jest szczególnie wysokie w przypadku instytucji finansowych w okresach dużej zmienności na rynkach. Co więcej, złożoność ΔCoVaR czyni tę miarę bardziej wrażliwą na parametryzację w porównaniu z MES, co sugeruje, że bardziej złożone MRS mogą prowadzić do większego ryzyka modelowego.

W badaniu wykorzystano symulację Monte Carlo do oszacowania MES i ΔCoVaR dla 47 banków europejskich. Ryzyko modelu skwantyfikowano za pomocą miary rozpiętości, która mierzy rozproszenie oszacowań MRS w przypadku różnych konfiguracji parametrów. Aby przeanalizować zmienność w rankingach banków (i ich koszykowaniu), użyto wskaźnika alfa Krippendorffa oraz procentowej zgodności. Cztery parametry podlegały zmianie w procesie MC: włączenie/wyłączenie zależności między bankami poprzez macierze korelacji, typ rozkładu (normalny lub t-Studenta), typ okna estymacji (ruchome lub ważone wykładniczo) oraz długość okresu historycznego (125, 250 lub 500 dni). Rozważano dwa punkty widzenia: (1) gdy rozkład MC (normaly/t-Studenta oraz korelacje) był traktowany jako parametr wraz z typem okna i okresem historycznym (jeden model z czterema parametrami, tj. 24 kombinacje); (2) gdy wybór rozkładu MC działał jako specyfikacja modelu i zmieniano tylko dwa parametry: typ okna oraz okres historyczny (cztery modele, każdy z dwoma parametrami, tj. sześć kombinacji dla każdego modelu). Ta metoda zapewnia efektywną obliczeniowo alternatywę dla metod bootstrapowych i umożliwia analizę wrażliwości przez klasyfikowanie wpływu poszczególnych parametrów na ryzyko modelu.

Wyniki wskazują na znaczne ryzyko modelu zarówno dla MES, jak i ΔCoVaR, z rozpiętościami wahającymi się od 68% do ponad 180% średnich oszacowań w różnych warunkach. Zaobserwowano również, że ryzyko modelu zwiększa się wraz z poziomem ufności. Alfa Krippendorffa dla zgodności między oceniającymi oparta na SRMs wahała się od 0,60 do 0,98 dla MES oraz od 0,03 do 0,74 dla ΔCoVaR, podczas gdy procentowa zgodność wynosiła 37–89% dla MES i 18–41% dla ΔCoVaR. Wyniki te pokazują dużą zmienność zgodności rankingów w odniesieniu do różnych kombinacji parametrów.

ΔCoVaR wykazuje większe ryzyko modelu niż MES, co potwierdza, że wzrost złożoności MRS zwiększa niepewność. Głównym parametrem przyczyniającym się do ryzyka modelu była długość okresu historycznego, co wskazuje, że pamięć o przeszłych wydarzeniach najsilniej wpływa na zmienność wyników MRS. Typ rozkładu również odgrywał istotną rolę, podkreślał bowiem znaczenie modelowania ogonów rozkładu w MRS. Wyniki te sugerują, że różne konfiguracje parametrów mogą prowadzić do zupełnie odmiennych rankingów systemowo ważnych instytucji finansowych, komplikując decyzje regulacyjne.

W pracy podkreślono znaczenie uwzględnienia ryzyka modelu w pomiarze ryzyka systemowego. Znaczna rozbieżność oszacowań MRS, spowodowana zmiennością parametrów, świadczy o potrzebie ostrożnej interpretacji tych modeli w kontekstach regulacyjnych. Regulatorzy powinni rozważyć podawanie zakresu oszacowań MRS (np. scenariusze pesymistyczne, optymistyczne i pośrednie), aby lepiej ocenić stabilność systemu finansowego, ponieważ wyniki wskazują, że ryzyko modelu wpływa na ranking banków. Przedstawiona tutaj metodyka może być wykorzystana w innych modelach parametrycznych i stanowi ramy analizy wrażliwości miar ryzyka na parametry.

Jednym z ograniczeń tego badania jest skupienie się wyłącznie na ryzyku parametryzacji. Inne źródła ryzyka modelu, takie jak jakość danych i problemy z implementacją, nie zostały zbadane w niniejszej pracy. Dodatkowo, chociaż miara rozpiętości skutecznie podkreśla ryzyko modelu, nie obejmuje w pełni wszystkich aspektów niepewności, takich jak interakcje między parametrami. Przyszłe badania powinny zbadać te wymiary, aby zapewnić bardziej kompleksowe zrozumienie wiarygodności MRS.

Słowa kluczowe: ryzyko modelu, ryzyko systemowe, Marginal Expected Shortfall, Delta Conditional Value at Risk, Monte Carlo